BVRIT HYDERABAD college of engineering for women
Employee Termination

DEPARTMENT OF CSE(AI & ML)

Team No: 3

20WH1A6611 - D.Joanna
20WH1A6612 - R.Pranathi
20WH1A6614 - A.Chithra Bhanu
20WH1A6615 - A.Tejaswini
20WH1A6616 - H.Rupa Kiranmai

April 1, 2023

PROBLEM STATEMENT

- Termination is one of the activities which happens in any organization. There are several factors which are considered for terminating an employee from his job. HR records all the information and facts about all employees.
- The challenge is to identify whom to terminate and keep the productivity of the company in an increasing order.

PYTHON PACKAGES AND LIBRARIES

- Numpy
- Pandas
- Matplotlib
- Sklearn
- tkinter
- pickle
- gradio

Types of Algorithms

- K-Nearest Neighbour
- Support Vector Machine
- Random Forest Classifier
- Logistic Regression
- Decision Tree
- Navie Bayes Classifier

k-Nearest Neighbour

- K-Nearest Neighbour is one of the simplest Machine Learning algorithms based on Supervised Learning technique.

- K-NN algorithm assumes the similarity between the new case/data and available cases and put the new case into the category that is most similar to the available categories

- K-Nearest Neighbors Accuracy: 77.78

Support Vector Classifier

- Support vector Machine algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future.

- SVM algorithm can be used for Face detection, image classification, text categorization.

- Support Vector Machine Accuracy: 92.06

Random Forest Classifier

- Random Forest is a process of combining multiple classifiers to solve a complex problem and to improve the performance of the model.

- The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.

- Random Forest Accuracy: 90.48

Logistic Regression Classifier

- Logistic regression is used for predicting the categorical dependent variable using a given set of independent variables.

- Logistic Regression is a significant machine learning algorithm because it has the ability to provide probabilities and classify new data using continuous and discrete datasets.

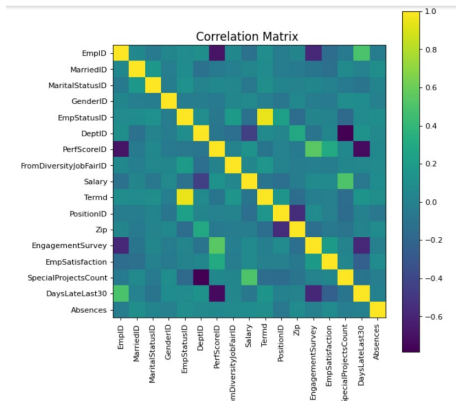- Logistic Regression Accuracy: 87.30

Decision tree

- Decision Tree is a Supervised learning technique that can be used for both classification and Regression problems, but mostly it is preferred for solving Classification problems.

- It is a tree-structured classifier, where internal nodes represent the features of a dataset, branches represent the decision rules and each leaf node represents the outcome..
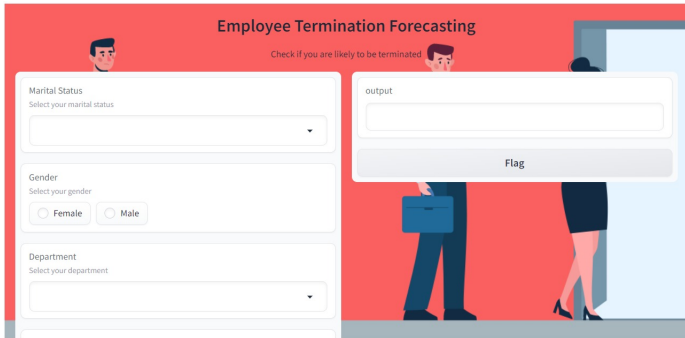
- Decision Tree Accuracy: 65.08

Naive Bayes Classifier

- Naïve Bayes algorithm is a supervised learning algorithm, which is based on Bayes theorem and used for solving classification problems.

- It is mainly used in text classification that includes a high-dimensional training dataset.

- Naive Bayes Accuracy: 79.37

Correlation Matrix

Figure: Output

Comparison Table

	MLA used	Train Accuracy	Test Accuracy	Precission	Recall	AUC
2	SVC	0.9476	0.9206	1.000000	0.772727	0.886364
5	RandomForestClassifier	0.9597	0.9048	1.000000	0.727273	0.863636
1	LogisticRegression	0.8710	0.8730	1.000000	0.636364	0.818182
4	MultinomialNB	0.8024	0.7937	0.800000	0.545455	0.736142
0	KNeighborsClassifier	0.8468	0.7778	0.722222	0.590909	0.734479
3	DecisionTreeClassifier	0.6694	0.6508	0.000000	0.000000	0.500000

Figure: Models Comparison

THANK YOU